Access Link Bottleneck Detection
Algorithm Description and Validation

Srikanth Sundaresan, Nick Feamster

WTF can compute the bandwidth of flows by capturing all flows at the access point, yet it is more difficult to identify whether the achieved throughput is throttled by a capacity constrained link in the path, or by other factors such as loss, or the application generating the traffic.

Intuition: Bottlenecks smooth packet arrival rates. Our first detector uses the coefficient of variation of packet interarrival times to determine whether the access link is introducing a performance bottleneck in the home network. Our detector is based on the following intuition: Because a bottleneck link services packets at a rate slower than they arrive, queues build up at the link, and the link thus paces at a relatively even rate. Packets upstream of the bottleneck will arrive according to the natural variation induced by TCP congestion control, but downstream of the bottleneck link, packets will be more evenly spaced.

Figures 1a and 1b illustrate this effect. The figures show the throughput of a TCP connection at the WAN port of the access point, at 10 ms granularity. In Figure 1a, the access link is at 100 Mbps, and the wireless link is 802.11a, with a clean channel. The wireless link is obviously the bottleneck here, as the maximum TCP rate it can support is less than 100 Mbps (about 21 Mbps in this case). We see the variation in instantaneous throughput effect caused by congestion control. Figure 1b shows the case where the access link is the bottleneck. In this case, the access link is shaped to 3 Mbps, while the wireless link is not modified. We see from the throughput plots that, as expected, the wireless is not the bottleneck, and there is very little variation in the throughput.

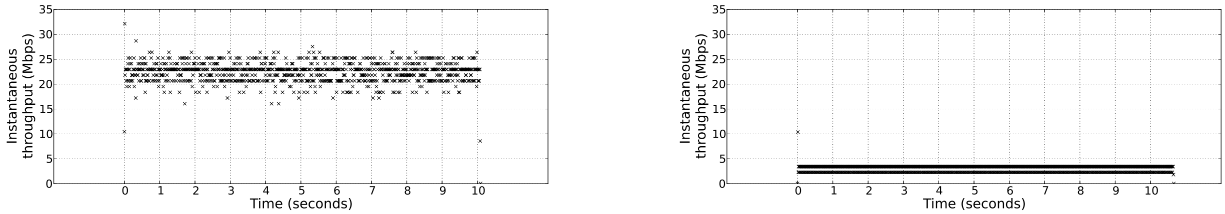
Informally, we expect to see high variance in the packet inter-arrival times before the bottleneck link due to congestion control, but significantly lower variance after the bottleneck link itself because of the buffer's smoothing effect. To identify whether the access link is shaping traffic, WTF determines whether it is more likely or not that the access link is the bottleneck, based on the observed values of the packet interarrival time, cv_t. We formalize this notion below, in the form of a maximum likelihood estimator.

Based on this intuition, we design a maximum likelihood detector that uses the coefficient of variation of packet inter-arrival times on the WAN side of the access point, cv_t, to determine whether the access link is the bottleneck. Our detector is based on a decision rule that determines whether the "access link bottleneck" event, B, occurs given a particular observed value of cv_t during a particular time period. We first compute the conditional probabilities $f_{cv_t|B}(cv_t|B)$ and $f_{cv_t|\overline{B}}(cv_t|\overline{B})$ in our controlled setting, where we use our ability to control the throughput of the upstream link to introduce a bottleneck on the access link. We then define our decision rule in terms of the likelihood ratio:

$$\Lambda(r) = \frac{f_{cv_t|B}(cv_t = v|B)}{f_{cv_t|\overline{B}}(cv_t = v|\overline{B})}$$

where cv_t is the coefficient of variation of packet interarrival time for packets over the observation window. $\Lambda(r) > \gamma$, the detector says that the access link is the bottleneck (*i.e.*, it is more likely than not, given the observation of cv_t, that the prior is the event B). We can tune the detector by varying the value of the threshold, γ; higher values of γ will result in a higher detection rate, but also a higher false positive. Given $\Lambda(r)$, we can thus determine the probability of a false positive and the probability of detection for different values of γ. These ranges of false positives and detection are commonly known as a receiver operating characteristic (ROC) for a decision rule.

Validation. We first evaluate the detection accuracy of the algorithm in this setting for different values of the maximum likelihood threshold, γ. Figure 2 shows the receiver operating characteristic for this detector. When the threshold



(a) *Wireless link bottleneck. Instantaneous throughput at the wide-area interface varies at short time scales due to high variance in packet inter-arrival times.*

(b) *Access link bottleneck. Instantaneous throughput at the wide-area interface is steady, due relatively uniform packet inter-arrival times caused by upstream shaping.*

Figure 1: *Behavior of packet inter-arrival times when the access-link is the bottleneck, and when it is not.*

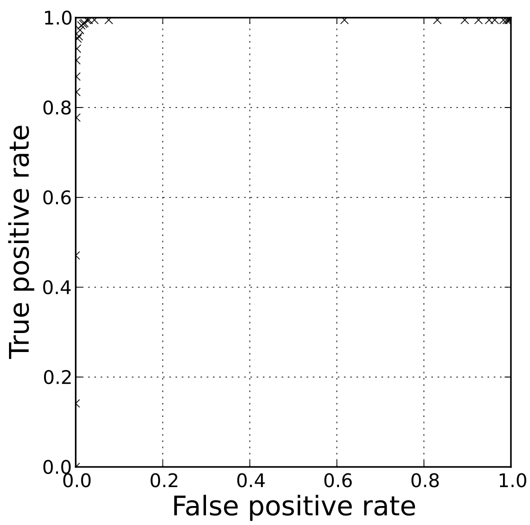


Figure 2: *Receiver operating characteristic for access link bottleneck detection using the coefficient of variation of packet interarrival time.*

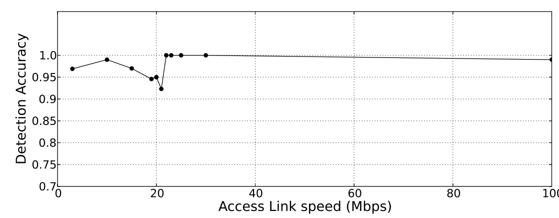


Figure 3: *Detection accuracy with the wireless channel held constant at about 21 Mbps. Detection accuracy is high except when the access link speed is similar to the wireless speed.*

is low (close to zero), it will always identify the wireless as the bottleneck, and when it is too high, it will always identify the access link as the bottleneck. Fortunately, the detector shows a high detection rate and low false positive rate for a wide range of values. The results indicate that detection accuracy remains high for a wide range of threshold settings. For the purposes of this paper, we use a threshold of $cv_t < 0.8$ to declarare the access link the bottleneck.

We also explore detection accuracy under a variety of downstream access link throughputs. Figure 3 shows the results of this experiment. We see that detection accuracy is generally quite high, in excess of 90% in all cases. The threshold value used here is 0.8. Detection accuracy drops when the access link is shaped at around 21 Mbps which is approximately the maximum speed of the wireless speed. In such cases, the capacity of the access link is close enough to the wireless capacity that the wireless could in fact be the bottleneck in some cases. As the wireless and the access link speeds converge, it may not even be meaningful to try label one side as the bottleneck due to natural variations. We see, however, that as the difference between the two sides increases, the detection accuracy is very close to 100%.